WAVELETS AND FIELD FORECAST VERIFICATION

DRAFT

Cornell University Biometrics Unit Technical Report

BU 1318M

17 February 1996

William M. Briggs wmb26@cornell.edu

Richard A. Levine ral4@cornell.edu

Cornell University

1. INTRODUCTION

The use of wavelets in data analysis is a rapidly growing field that is only recently being exploited in the atmospheric sciences. The beauty of wavelet analysis is frequency *and* time decomposition of data (e.g. time series data) is possible. In contrast, traditional Fourier analysis supplies only frequency decomposition. Meteorologists may be most familiar with wavelets in turbulence studies (e.g. Gao and Li, 1993; Hagelberg and Gamage, 1994; Katul and Vidakovic, 1995), time series analysis (e.g. Lau and Weng, 1995; Weng and Lau, 1994; Meyers et al., 1993), and other miscellaneous geophysical data analyses (e.g. Chao and Naito, 1995; Kumar and Foufoula-Georgiou, 1993; Serrano et al., 1992). One of the largest (non-meteorological) applications of wavelets has been in the field of image analysis (Laurent et al., 1993), to remove noise or compress images for data transmission (see Press et al., 1992, for a good introduction). Excellent mathematical introductions to the theory of wavelets may be found in Chui et al. (1994) and Daubechies (1992). An introduction with examples catering to geoscientists has been presented by Kumar and Foufoula-Georgiou (1994). We additionally recommend the very readable paper by Vidakovic and Müller (1994), non-threateningly entitled "Wavelets for Kids."

Wavelets have also found their way into statistics, particularly in non-parametric regression (e.g. Nason, 1994) and non-parametric function and density estimation (e.g. Donoho and Johnstone, 1994; Donoho et al., 1995). It is with these recent statistical insights that we hope to demonstrate how wavelet analysis can be used in providing measures to compare images and scores to assist in field forecasts verification. The

measure of closeness between images is a general problem which is encountered in areas other than meteorology; for example, in the quantification of electronic transmission error, image compression algorithms, or pattern recognition studies. We also remark on how statistical techniques using wavelets can be incorporated into other data analysis settings, e.g. univariate time series studies. In this article we use the terms image, grid, and field interchangeably.

The primary purpose of field comparison for meteorologists is in forecast diagnostics. Typically, we are presented with an analysis field, which is given as "truth," and a corresponding forecast field . Our goal is to rate how "close" the forecast field is to the analysis field. This "closeness" gives us a sense of the predictive power of the forecast model that generated the field forecast. The question of declaring meteorological analysis fields as "truth" (see Daley, 1993, for a discussion) is not dealt with here except to mention that the techniques for generating analysis fields do contain error.

Various objective statistical measures have been developed to measure closeness. Among these are the anomaly correlation coefficient (ACC, Miyakoda et al., 1972), the familiar Pearson product moment correlation, and the root mean square error (RMSE) and its variants (e.g. normalized RMSE, mean absolute deviation, etc.). However, these measures have all been judged inadequate. Murphy and Epstein (1989) and Radok and Brown (1993) detail the statistical problems with the ACC (e.g. the well known bias towards a value of 0.5). We should also note the ACC cannot even be computed when there is no climatological field available, a non-trivial concern in measuring image closeness. Taylor (1991) shows that other measures in use (RMSE and mean-absolute deviation) are also

poor, in the sense of giving incomplete information, when used in image comparison.

One of the major difficulties with these traditional image comparison/field closeness measures is that they are single numbers (point or univariate scores). Undoubtedly the application of point scores is popular due to the ease and quickness of use. But the goal of these scores is to compare the closeness of two complex fields. Any one number which attempts to squeeze all the information about intricate spatial and distributional aspects of the closeness will almost certainly be inadequate (Murphy and Winkler, 1987; Murphy, 1991). Section 2 will discuss this issue in more detail.

We recognize, however, that the computational simplicity and the easy grasp of point scores are appealing in practice. Therefore, we will give a method which uses wavelets for "cleaning up," in the sense of removing noise, some common univariate scores. This idea is analogous to techniques that have been used in the past, e.g. spectrally filtering fields (a discussion appears in Van Den Dool and Rukhovets, 1994) or filtering fields with empirical orthogonal functions (EOFs; e.g. Livezey et al., 1995). Once two fields have been filtered, either to remove noise or to remove data at wavelengths of no interest, point scores can be recomputed to give insight as to closeness of the filtered fields. Wavelets can also be utilized to filter fields, and new research shows how this may be done in a statistically optimal manner.

A complete distributional description of closeness is unlikely, unless significant simplifying assumptions are made, due to the complexity and dimensionality of the problem (Murphy, 1991). We can do better than point scores, however. For example, Hoffman et al. (1995) present a

multivariate score based on what they term the distortion representation of forecast error. Their approach "breaks" the score into pieces that explain different aspects of the error for use in a specific application. As we will see, the wavelet approach is similar in that wavelet transformations decompose fields into scales representing different detail levels of the field. We exploit this characteristic to develop intuitive multivariate measures of closeness to compare the forecast and analysis field at the multiple scales of detail.

The paper unfolds as follows. In Section 2 we discuss the traditional approach of field forecast verification. We provide further impetus for improving closeness measures via wavelets. In order to describe the application of wavelets to forecast verification, we familiarize the reader with wavelet mathematics. Therefore, we first delve, briefly, into the theory of discrete wavelet transforms in Section 3. We also present recent advances in wavelet statistics which will be used in the latter part of the paper. We emphasize that these statistical methods are applicable in any data analytic setting. In Section 4 we show how wavelets can be used in statistical image comparison. We also carry out simulations using actual and synthetic (but realistic) forecast fields and (an actual) analysis field to illustrate the filtering of univariate closeness measures. We then construct multivariate closeness measures based on wavelet decomposition. Section 5 contains concluding remarks.

2. TRADITIONAL STATISTICAL APPROACHES

There are two problems in field forecast verification. The first is to quantify or define what is meant by verification and the second is to

develop and understand scores based on this quantification. We first outline the goal of absolute field forecast verification and then examine what common point scores are actually measuring in the field forecast framework.

The use of a single measure (point score) in evaluating any forecast, whether for a field or for a univariate variable (e.g. probability of precipitation forecast), is troublesome because no single measure can capture all relevant information regarding accuracy. Murphy and Winkler (1987) and Murphy (1991) discuss this consequence as the problem of absolute forecast verification. They show that any scoring system must include enough information to allow for reconstruction of the joint distribution of the forecast and analysis values. The familiar example of (point location) probability of precipitation (PoP) forecasts highlights this fact. In scoring PoP forecasts we typically examine both calibration and sharpness (reliability). Neither measure alone is sufficient in fully explaining the overall performance of the forecasts. Likewise, we cannot hope to reconstruct the complex spatial information in field forecasts from a point score. We need a better measure of the joint distribution of the (usually non probabilistic) forecast and analysis fields. However, to accomplish this task, we must find a way to overcome the complexity and overwhelming dimensionality of the problem.

Following Murphy (1991), we define the dimensionality of forecast verification to be the minimum number of specifications needed to fully describe the joint distribution of forecasts and observations. Numerically, dimensionality is defined as the number of distinct forecasts times the number of distinct observations minus one. The dimensionality of the PoP verification example, when there are 11 distinct probabilities allowed (e.g.

0.0, 0.1, 0.2, etc.), is 21. For field forecasts, this number is essentially the number of grid points squared. The dimensionality of forecast verification for even a small grid is huge (since fields contain continuous variables, parametric models for the variables under consideration can be introduced to reduce the dimensionality). In this paper we are not presenting a complete solution to the absolute field forecast verification problem. Instead, we hope to persuade the reader that reliance on single scores is misguided and that some form of a multivalued score is better. Later, we will show that an objective multivariate measure can be developed using wavelets.

As mentioned in Section 1, common point scores used in field forecast verification are the ACC (resembling the ordinary sample linear correlation, r) and RMSE. In using these (or any) scores it is important to understand what they are attempting to measure.

It is convenient to present the ACC, in matrix norm form, as

$$\frac{\sum diag\ (\mathbf{A}-\mathbf{C})^{\mathrm{T}}(\mathbf{F}-\mathbf{C})}{\|\mathbf{A}-\mathbf{C}\|_f \|\mathbf{F}-\mathbf{C}\|_f}, \tag{2.1}$$

where $\mathbf{A}$ is the analysis field, $\mathbf{F}$ is the forecast field, $\mathbf{C}$ is the climate field, $\|\cdot\|_f$ indicates the Frobenius norm of the argument, and the sum in the numerator is over the diagonal elements of the corresponding matrix (Golub and Van Loan, 1989). This score can only be used when $\mathbf{C}$ is meaningfully defined, hence it is constrained to meteorological-type fields. Writing the ACC in this form emphasizes the fact that we are examining a relationship between two matrices *not* simply a collection of independent observations, i.e. the values at each grid point cannot be statistically considered independently. The sampling distribution of the ACC is unknown. However, it is well known that the measure is biased, under

general conditions, towards the value 0.5 (Murphy and Epstein, 1989; Radok and Brown, 1993). More on the ACC will be said in a moment.

It is helpful to examine common closeness scores for statistical problems in general to better grasp what these measures are doing when extended to fields. Given two arbitrary sets of data, a_i and f_i, i = 1,...n, each set independent and identically distributed (iid) realizations of random variables of mean zero, we can compute the

$$\text{RMSE} = \|\mathbf{a} - \mathbf{f}\| / \sqrt{n} \tag{2.2}$$

between them, and ordinary correlation

$$r = \sum \mathbf{a}^T \mathbf{f} / \|\mathbf{a}\| \cdot \|\mathbf{f}\| \tag{2.3}$$

between them, where $\mathbf{a}$ and $\mathbf{f}$ are the data in vector form and the norms correspond to the Euclidean norm. These scores have an interesting geometric interpretation: these are the distance between two n-dimensional vectors and the cosine of the angle θ between them, respectively (see Figure 1). Murphy and Epstein (1989) present the following interpretation of MSE,

$$MSE = BIAS^2 + s_a^2 + s_f^2 - 2 s_a s_f r, \tag{2.4}$$

where $s_\bullet^2$ is the variance of the respective data vector. Equation (2.4) also has a well known geometric interpretation (and without loss of generality we ignore bias which is important only when the means of $\mathbf{a}$ and $\mathbf{f}$ is not equal to zero),

$$\|\mathbf{a} - \mathbf{f}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{f}\|^2 - 2\|\mathbf{a}\|\|\mathbf{f}\|\cos(\theta), \tag{2.5}$$

where each term in (2.5) is synonymous with each respective term in (2.4) (except for bias). Equation (2.5) is known as the Law of Cosines describing the relationship between the squared distance and angle between two vectors (see Fig. 1).

In scoring closeness of typical field forecasts, **A** and **F** are used in place of **a** and **f**, and then treated like *vectors* and not *matrices*. This substitution is technically incorrect as the key assumption of independence is not met and all spatial information about closeness is removed. This means that the resulting closeness scores cannot be interpreted in the traditional manner. We argue that, instead of the Frobenius vector norm commonly used in (2.2) and (2.3), the more proper L_2 matrix norm should be used. The L_2 matrix norm is commonly used as a measure of distance on the space defined (or generated) by the matrices at hand (Golub, and Van Loan, 1989). Examples in Section 4 will demonstrate this.

As mentioned, if two sets of random variables are iid, then exact forms of the sampling distributions of the closeness scores (2.2) and (2.3) are known. We can then use these sampling distributions to assess the statistical significance to any particular score received between two sets of data. The problem with field forecasts is that the iid assumption is violated because meteorological fields have complex covariance structures between different spatial grid points (e.g. Perrie and Toulany, 1989). Adjacent, and even far-removed, grid points are highly correlated with one another. Therefore, in principal, we must treat the field forecast as one sample from an n-dimensional multivariate distribution, where n is the number of grid points. This assumption is in contrast to n samples from a univariate distribution as is commonly assumed. There exist multivariate techniques for statistically comparing two sets of variables (or two fields) with complex correlation patterns. However these standard multivariate approaches fail in our case due to the lack of replications in the observations (e.g. see Anderson, 1984, chapter 10; recall, we are trying to develop a closeness measure for a single forecast and analysis field). If we

could somehow remove the covariance between observations on our grid, i.e. make the grid points statistically independent, we could use and correctly interpret the standard scores. In a certain sense, discrete wavelets provide a method to orthogonal transform (i.e. make statistically independent) data. This transform allows us to partition the original data in an objective manner. We can then compute standard scores for each of these partitions resulting in a multivalued score for field forecasts.

We close this section with a few remarks on the ACC. Note that ACC (2.1) resembles r (2.3), except for the subtraction of $\mathbf{C}$ with, as Wilks (1995) notes, the ACC equaling r when $\sum A_i - C_i = \sum F_i - C_i = 0$, and the forecast is as variable as the analysis. Thus, for large fields we can write ACC $\approx r$. This approximation allows us to offer a cute geometric interpretation of why the ACC (actually r) is biased towards a value of 0.5. It can be shown that if two vectors (of length k) were "dropped" randomly on the real line, such that the distribution of the angles θ_A and θ_F between the $k \times 1$ vector $[1 \quad 0 \quad \cdots \quad 0]$ (i.e. the kth dimensional "x-axis") and the vectors $\mathbf{a}$ and $\mathbf{f}$ are distributed as Uniform $[0, \pi]$, the expected value of the cosine of the angle θ between $\mathbf{a}$ and $\mathbf{f}$, i.e. r, is equal to 0.5. Constraining the vectors to lie in the interval $[0, \pi]$ simulates the condition that the forecast has a correlation between -1.0 and 1.0 (see Murphy and Epstein, 1989 for the conditions necessary on fields that produce the ACC bias of 0.5). If we allow the two vectors to be placed anywhere in the interval $[-\pi, \pi]$ the expected value of the cosine of the angle between them, r, is equal to 0.0 as expected. So very simply, we would expect that, on average and even for poor forecasts, that ACC $\approx$ 0.5. This makes interpretation tricky as the bias must always be considered in the closeness rating process.

See Person, A., 1996: Forecast Error and Inconsistency. Preprints, 13th Conf. on WX Prob. Proceedings: Also Person: The relation between FCST Inconsistency and skill in deterministic and ensemble prediction

Various authors have sought methods in which to maximize the standard scores, namely the ACC, for a particular forecast-analysis field pair in the sense of examining the score for only a part of the forecast. Typically this has been done by spectral filtering (e.g. Branstator et al., 1993; Van Den Dool and Rukhovets, 1994) or through empirical orthogonal functions (EOFs; e.g. Livezey et al., 1995). In the first method, wavelengths considered "too small" are filtered out of both the analysis and forecast fields and the standard scores are recomputed. This filtering tends to eliminate small scale variation so that the filtered ACC is higher and the filtered RMSE is lower. This result is also true for EOF filtering. Both analysis and forecast fields are transformed in the standard way and only the EOFs above a predetermined significance level are kept, the rest are set to zero. Again, small scale variation from the fields is removed and the standard scores improve (if only one unique forecast-analysis pair is available, singular value decomposition, instead of EOF decomposition, can be performed with only the "largest" eigenvalues corresponding to a predetermined percentage of variance). Wavelet transforms are an improvement over EOF filtering because interpretation of wavelet scales is still physically justifiable while interpretation of EOFs other than the first are not strictly interpretable (without, say, rotation which eliminates the nice properties of orthogonality; e.g. Preisendorfer, 1988).

Both ideas, EOFs and wavelets, have merit and can be used for intensive field forecast diagnostics. Examination of forecast performance by wavelength via spectral analysis can point to certain model flaws needing correction, or to suggest reliability of certain scales of forecasts (Branstator et al., 1993). It is possible to partition forecast error into

orthogonal blocks by EOF provided a history of forecasts and analyses are available. There exist formal statistics that allow one to identify the noise floor in EOF analysis (see Wilks, 1995), but is not clear from the literature that these methods are used with any consistency. More often, an arbitrary cut-off is often chosen, say those EOFs that explain at least 90% of the variance.

Discrete wavelet transforms allow us to objectively identify and eliminate insignificant contributions to closeness scores. It will be shown, through simulations and examples, that the standard scores tend to improve when using wavelet transforms.

3. THE DISCRETE WAVELET TRANSFORM

This section will introduce the terminology and algorithms used in performing the discrete wavelet transform, but will skip detailed mathematical theory and justification. For more specifics see the sources mentioned in Section 1. We use discrete as opposed to continuous wavelet transforms because of the orthogonality property of the former. This quality ensures that the wavelet coefficients (to be defined) are statistically independent from one scale to the next. More common continuous wavelets transforms do not provide the same statistical advantages.

We first outline the mechanics of the discrete transform for images. The methodology developed is quite general and may be used for either one- or two-dimensional data. In order to transform, we must first select a so-called mother wavelet. We present statistical methods based on entropy principles that choose this mother wavelet from a library of potential choices in an objective manner. After the data is transformed into the

wavelet space, we can eliminate insignificant elements by a process known as thresholding. We detail recent advances in wavelet statistics that show how to threshold in an objective manner. Thresholding can also be done on one- or two-dimensional data. For example, thresholding can be employed in nonparametric regression problems (Nason, 1994).

3.1 *Mechanics of the discrete transform* $n_r \times n_c$

Let an image (or grid, or field) be written as the real-valued matrix $\mathbf{G}$ consisting of n_r rows and n_c columns. It is possible to construct an orthogonal operator Ψ, called the mother wavelet, such that the discrete wavelet transform (WT) is given by

$$\mathcal{W} = \Psi \mathbf{G}, \tag{3.1}$$

where $\mathcal{W}$ is the WT transform matrix of $\mathbf{G}$ whose elements $d_{i,j}$ are called wavelet coefficients (Donoho and Johnstone, 1994). In other words, we project $\mathbf{G}$ onto the orthonormal basis Ψ yielding a transformed field $\mathcal{W}$. A discussion of how the elements of Ψ are found such that it is an orthogonal basis can be found in, for example, Daubechies (1992). Certainly Ψ is not unique, and finding new representations for this matrix is an active area of research (e.g. Chui et al., 1994). If Ψ is well chosen, the transformed field or matrix $\mathcal{W}$ can be sparse, i.e. many, even the vast majority, of elements may be at or near zero, with only a few elements being relatively large. This characteristic has allowed WTs to be used in image compression with great success (Laurent et al., 1993). These small or zero elements can be eliminated by thresholding and thus less information must be stored. Wavelets have the ability to represent $\mathbf{G}$ by retaining only a fraction of the original number of data values.

Equation (3.1) is readily invertible, i.e. $\Psi^{-1}\mathcal{W} = \mathbf{G}$. Typically (3.1) is only performed for matrices of the size $n_r = 2^{k+1}$ and $n_c = 2^{l+1}$ (Donoho and Johnstone, 1994; Press et al., 1992). These dyadic or "power of two" matrices will be used in the present work. Specifying matrices in this form greatly aids in reducing the number of calculations necessary to perform the transform (of order $O(n)$, Vidakovic and Müller, 1994). It is possible to form a dyadic matrix from any non-dyadic matrix simply by "padding" the end rows and columns with dummy values, say, zeros. If this is done, however, it must be kept in mind when performing any subsequent analyses.

The approach taken here for wavelet transforming an image is the sequential method (Saito, 1994) as outlined in Press et al. (1992). We first WT each row of $\mathbf{G}$, and then WT the columns of the transformed rows to produce one matrix of wavelet coefficients. Another method to WT an image breaks the original image matrix into separate detail and "mean" matrices, all of less dimension than the original matrix (Saito, 1994; Daubechies, 1992; Vidakovic and Müller, 1994). In our context we prefer the first method because we desire to work with the block scale coefficients which are more easily interpretable. Both are equivalent in the mathematical sense.

To be more explicit, $\mathcal{W}$ will contain $l+1$ (from $n_c = 2^{l+1}$) orthogonal blocks. Each block has $n_r = 2^{k+1}$ rows. The number of columns in the first block is 2. The number of columns in the remaining l blocks is m^2, $m = 1,2,\ldots,l$. As can be seen, the number of columns in each successive block increases by a power of two indicating wavelet representation at finer and finer scales of the original data. The first block of two columns corresponds to, what may be thought of, as the "mean" of

the data with respect to the structure of Ψ. The second block cuts the data in half and the columns represent how each of these halves resemble the structure of Ψ. This interpretation can be, admittedly, confusing at first glance but we urge readers to stop and appreciate this process as we use the final WT matrix $\mathcal{W}$ in the following sections.

As an example we present Figure 3.1. Fig. 3.1a is an $n_r = 2^{3+1}$=16 by $n_c = 2^{4+1}$=32 field of 500 MB heights over (roughly) North America (9 DEC 1992, 12Z; data taken from the National Center for Atmospheric Research's archive of a European Center for Medium range Weather Forecasting analysis from the World Climate Research Programme). A background map has been left out of the figure to emphasize detail (and to reinforce the notion we are looking an image). Fig. 3.1b is the contoured WT matrix using a Daubechies 8 (sometimes called a "daublet") mother wavelet of the same image. Contouring is only one way to view this matrix. Alternatively, we could look directly at the matrix values. Notice that only a few coefficients are relatively larger than zero: large coefficients are indicated by the tight contour gradients. The remaining smaller coefficients may be insignificant, i.e. noise. In Section 3.3 we show how we can safely ignore these potentially negligible coefficients.

3.2 *Choice of the Mother Wavelet*

There exist many different mother wavelets. The question of which mother wavelet is best to use for transforming a particular field or vector of data is an important and sometimes neglected one. The ideal WT reduces the data to the greatest degree, i.e. produces the most coefficients near zero. To aid in this choice we will study the L_2 matrix norm of the WT matrix $\|\mathcal{W}\|_2$ (Vidakovic and Müller, 1994). The L_2 matrix norm (as

opposed to the Frobenius or Euclidean norm) norm is the most commonly accepted matrix norm (Golub and Van Loan, 1989). In this work, as in much of the current statistical literature, we also define the norm as the total "energy" of the matrix. The procedures developed in this section are equally applicable to one-dimensional data sets (e.g. time series, or regression problems).

We adopt the procedure developed by Goel and Vidakovic (1994) and Katul and Vidakovic (1994) used in choosing the best mother wavelet. It is based on minimizing the entropy of the wavelet transformed matrix. The idea is that the WT disbalances the energy of an image and that the most disbalanced transform is best. This approach makes intuitive sense in that the minimum entropy will be with the transform that produces the greatest ratio of few large coefficients to many small coefficients. We start with a library of mother wavelets, compute an entropy score for each transform, and pick the mother wavelet which produces the best score. We will have more to say on the subject of this library in a moment, but for now we use the library of mother wavelets based on the increasing coefficients of Daubechies and Symmlet wavelets along with the Haar wavelet (Daubechies, 1992; Press et al., 1992).

The measure found to be the most resilient by Goel and Vidakovic (1994) is the Shannon entropy measure given by

$$\phi(\mathbf{d}) = -\sum_{i,j} d'_{i,j} \log d'_{i,j}, \tag{3.2}$$

where d'_i are the non-negative normalized wavelet coefficients, i.e. $d'_{i,j} = |d_{i,j}| / \sum |d_{i,j}|$ and $0 \log 0 = 0$ by definition. A similar argument was used by Saito (1994) in his algorithm to select a best basis from a library while simultaneously thresholding (thresholding will be discussed below).

The best WT will minimize $\phi(\mathbf{d})$ from the library of possible mother wavelets.

Recall there are a very large number of orthogonal bases Ψ. In order to make practical use of Shannon's entropy measure, we must limit ourselves to a finite few whose value is proven by experiment. Clearly, work needs to continue in this area. But whatever library is eventually selected for a task, the user must make explicit the eventual mother wavelet choice. We have had good success with both the Haar and the Daubechies mother wavelets and recommend these as starting points (although there are many others, such as Symmlets, Coiflets, etc.).

An example of Shannon's entropy measure applied to our library of mother wavelets is found in Table 1. Here we present the Haar wavelet and an increasing series of Daubechies wavelets with the corresponding entropy scores. Table 1 indicates that the Daubechies 8 mother wavelet is the optimal choice from our library.

Table 1. Comparing Entropy scores for various mother wavelets and the image of Fig. 1.

	Haar	Db 4	Db 6	Db 8	Db 10	Db 12	Db 16	Db 20	Db 30
Score	1.947	1.874	1.861	1.803	1.892	1.857	1.847	1.825	1.920

3.3 *Wavelet Thresholding*

Thresholding refers to the process of shrinking the coefficients of $\mathcal{W}$, i.e. setting to zero or shrinking towards zero certain coefficients, in an effort to remove insignificant information. Generally, there are two manners of thresholding, hard and soft. Hard thresholding is of the form

$$d_{i,j} = \begin{cases} 0 & |d_{i,j}| < \lambda \\ d_{i,j} & |d_{i,j}| \geq \lambda, \end{cases} \tag{3.3}$$

where all $|d_{i,j}| < \lambda$ are set to zero and the rest are kept. Soft thresholding is of the form

$$d_{i,j} = sign(d_{i,j})(|d_{i,j}| - \lambda)_+, \tag{3.4}$$

where all $|d_{i,j}| < \lambda$ are set to zero and the rest are shrunk toward zero by an amount λ (Donoho et al., 1995). Donoho and Johnstone (1994) maintain that the largest scales wavelet coefficients should be left unthresholded regardless of their size. Experiments with meteorological fields suggest this is a good practice. Hard thresholding is ideal for data compression, while statistical arguments suggest soft thresholding be used in analysis settings.

The best choice for λ depends on the assumed model for the data. Donoho and Johnstone (1994) and Donoho (1992) provide a method to find λ for one-dimensional data. They first suppose data of the form $y_i = f(x_i) + e_i$ is observed, where y_i are the data, $f(x_i)$ is the function generating the data, and e_i is noise distributed as $N(0, \sigma^2)$. It is a remarkable property of the WT that it has no effect on noise (Nason, 1994), i.e. if a pure noise data set is presented to wavelet transform, the resulting transformed data set will be indistinguishable (in a statistical sense) from the original untransformed image. This fact should not be surprising as the entire purpose of the WT is to capture real processes in the data, not noise.

The above authors proved that, if this form of the data is correct, then the λ that is minimax (under certain regularity conditions) is of the form

$$\hat{\lambda}_U = \sqrt{2 \log n}\,\hat{\sigma}/\sqrt{n}, \tag{3.5}$$

where $\hat{\sigma}$ is also estimated from the data as the median absolute deviation of the wavelet coefficients at the smallest scale (Donoho et al., 1995; they also divide $\hat{\sigma}$ by 0.6745 which slightly increases $\hat{\lambda}_U$). This $\hat{\lambda}_U$ is known as the "universal" thresholder. We note that this estimator may be ideal for many one-dimensional data sets.

Johnstone and Silverman (1994) sought to generalize the universal thresholder $\hat{\lambda}_U$ by allowing for data with correlated noise, which will certainly be present in images. Their version of $\hat{\lambda}_U$ is very similar to the universal one and is given by

$$\hat{\lambda}_{Um} = \sqrt{2\log n}\,\hat{\sigma}_m/\sqrt{n}, \tag{3.6}$$

where a different $\hat{\lambda}_{Um}$ is computed for each scale or block m and $\hat{\sigma}_h$ is the standard deviation of the wavelet coefficients at that scale.

A final thresholding estimate is due to Goel and Vidakovic (1995) and is used in an atmospheric turbulence study by Katul and Vidakovic (1995). The methodology is based on the Lorentz curve (found commonly in economics) and is appealing since no distributional form is prescribed for the noise. The Lorentz curve threshold is derived from entropy arguments as was the Shannon entropy measure for mother wavelet selection. It attempts to find the place in the data where the signal first rises above the noise. It written as

$$\hat{\lambda}_E = \frac{\|\mathcal{W}\|_2}{\sqrt{n}}. \tag{3.7}$$

Since atmospheric data is certainly highly correlated, we propose a combination of (3.6) and (3.7) as a threshold

$$\hat{\lambda}_{Em} = \frac{\|\mathcal{W}_m\|_2}{\sqrt{n_m}}. \tag{3.8}$$

This threshold gives us the nice properties of the entropy threshold applied to each scale or block m, for all $m > 2$. (The first two blocks are not

thresholded because of statistical optimality considerations which are discussed in Donoho and Johnstone, 1995).

Figure 3.2 presents an example of wavelet compression. The solid contour is identical to the analysis field of Fig. 3.1. The dashed lines are the result of WTing the analysis field with the Daubechies 8 mother wavelet, applying hard threshold (3.8), and inverse WTing. Only 131 of the original 512 data points were kept after thresholding, for a reduction to 26% of the original. DeVore (1993) shows that, for some images, this can be improved to as little as 5%!

4. USE OF WAVELETS WITH COMMON SCORES

As mentioned in Section 2, we have two goals. The first is to present a method to objectively improve, in the sense of removing insignificant information from, the standard closeness scores between fields. We will do this by WTing the analysis and forecast fields, soft thresholding, inverse WTing, then recomputing the scores. For examples we will present scores based on simulations of realistic forecast fields using the analysis field described in Section 3. Furthermore, multiple pseudo-forecasts decreasing in prediction accuracy (or increasing in forecast *badness*) will be simulated, and the original and modified scores will be compared. We choose to study simulations over real forecast fields because we are able to specify the exact characteristics of the forecast field. Thus we have one less unknown to consider when examining the behavior of the WT scoring system.

The second and more important objective is to develop a multivariate closeness score that better captures information about field closeness. An

actual numerical forecast will be used to demonstrate how WTing can be employed to construct a multivariate closeness score. The components of this measure are the standard scores computed for each scale of the WTed data. We can thus obtain, for example, an indication of how each scale contributes to the overall RMSE, or we can compute an RMSE for each scale.

4.1 *Removing Noise From Point Scores*

In a sense, each measure of closeness between two fields contains a certain amount of information that is not helpful or is insignificant. This insignificant information is the result of random noise, for example, in both the analysis and forecast fields. This is the same as saying, for example, that we observe the fields $\mathbf{A} = \mathbf{A}' + \varepsilon$, and $\mathbf{F} = \mathbf{F}' + \eta$, where $\mathbf{A}'$ and $\mathbf{F}'$ are the true fields and ε and η are fields of random error. Using WTs we can objectively identify and remove (with soft thresholding), at each scale, the insignificant portions of the data that are solely attributable to this noisei, i.e. WTs give estimates of $\mathbf{A}'$ and $\mathbf{F}'$. The recomputed, or modified, closeness scores between the fields $\mathbf{A}'$ and $\mathbf{F}'$ are thus giving us better, or tighter, information. One can argue that this line of inquiry is not completely justifiable and we would tend to agree. But it is a first attempt at approximating information we might be able to acquire from the sampling distributions of the scores if they were known.

In order to simulate the pseudo-forecast field $\mathbf{F}$, a covariance matrix corresponding to the data must be specified. We assume here the forecast $\mathbf{F} \sim N(\mathbf{A}, s\Sigma)$, i.e. a normal field with mean $\mathbf{A}$ (the analysis field) and covariance $s\Sigma$ (a 512 x 512 matrix). The scale parameter s allows simulation of various levels of forecast badness. For example, s = 1, say,

allows simulation of forecasts that are "good". A value of s = 400, simulates forecasts that are considered "bad". This technique is arguably subjective, but visual inspection of these pseudo-forecasts at various values of s indicated the model we have chosen is adequate for an initial study of the performance of the scores. The matrix Σ is chosen such that correlations decrease (from grid point to grid point) by 0.9^{ρ}, where $\rho = d\left((i,j),(i',j')\right)$ is a function returning the Euclidean distance between the grid points (i,j) and (i',j'). Adjacent grid points are defined as being one unit apart, etc. This choice leads to a smooth decrease in the correlations as grid point distance increases. Other choices could be made for the function $d(\bullet)$, such as an exponential function (Daley, 1992, chapter 4). Still more complex relationships describing correlation could be utilized in future studies, such as that described in Handcock and Wallis (1994). The normal field choice, however, will serve as a good starting point.

As an example, Figure 4.1 shows one realization of $\mathbf{F}$ with s = 400. The influence of $\mathbf{A}$ can be readily seen, although differences due to $s\Sigma$ are also visible. We argue that, as a first cut, these differences are adequate to simulate "real" forecasts.

The simulation to demonstrate the removal of noise from point scores proceeds as follows. For each fixed s, simulate 50 pseudo-forecasts $\mathbf{F}$. For each of these $\mathbf{F}$, compute the ACC($\mathbf{A}$, $\mathbf{F}$) and RMSE($\mathbf{A}$, $\mathbf{F}$). Now WT $\mathbf{A}$ and $\mathbf{F}$, apply soft threshold (3.8), inverse WT both fields and recompute the ACC and RMSE. The resultant 50 scores, both before WTing and after are averaged. A new value of s is fixed and the process repeated.

Figure 4.2 presents the results for s ranging from 1 to 400. Fig. 4.2a shows the ACC, and Fig. 4.2b displays the RMSE. For each, the solid line is the standard score while the dotted line is the WT modified score. As can be clearly seen, both scores show improvement after WTing in that the ACC increased and the RMSE decreased. At a value of $s = 400$, perhaps not untypical for extended range forecasts, the ACC improved by about 5% while the RMSE improved by about 18%.

We reran the simulation, this time including an EOF analysis (actually SVD as we have unique fields). For each simulated **F** and **A** we performed the SVD and kept only the leading three eigenvalues (about 80-90% of the variance), and set the remaining ones to zero. The data was then transformed back and the standard scores calculated. With pseudo-forecasts of the type we are using, we found there to be almost no noticeable improvement in the score's values. Of course, a change in the function $d(\bullet)$, or other simulation assumptions, or the use of actual forecasts may lead to different results in the scores.

Overall, the WT modified scores behaved as expected in that, as more and more noise was added to the analysis, and the noise corrupted analysis was compared with its uncorrupted self, the modified scores adjusted more strongly. These simulations have nothing to say about any particular bad forecast, i.e. we would not necessarily expect a true forecast with an error variance $s = 400$ to show as drastic an improvement in the scores. This is because a true forecast will be apart (or not close) from the analysis in ways other than noise. Rather, we are taking a known field and seeing how modified closeness scores perform when the noise level is specified.

4.2 *Example of a Multivariate Score*

We now present a detailed example of a multivariate score based on statistics computed in the wavelet space. Recall that for our particular data set **A**, there are $l+1$ blocks representing different orthogonal scales of the data. It is possible to both compute the degree to which each scale/block contributes to the overall score(s) and to compute scores for each scale. For the former we compute the percent of each score explained by scale after thresholding. For the latter, we first inverse WT each scale (setting the others to zero) and then compute scores for these inverse WT fields. Here, we use hard threshold (3.8) on the data to present cleaner images, although we stress that it is not necessary that this be done. These concepts can be made clear with an example.

Figure 4.3 details a 36-hour forecast made by the National Center for Atmospheric Research (NCAR) Community Climate Model-2 (CCM-2) for the analysis field of Fig. 3.1. The ACC between the two fields is 0.974. The RMSE is 21.83 m. To illustrate the concepts of the previous section with a real forecast, after WTing, thresholding, and inverse WTing, the modified ACC is 0.979, and the modified RMSE is 18.23 m. This represents a change of 0.6% and 4.9% respectively.

The percent each scale contributes to the ACC can be calculated by the following

$$\%ACC_h = \frac{\sum diag\left(A_h - C_h\right)^T \left(F_h - C_h\right)}{\sum diag(\mathbf{A}-\mathbf{C})^T(\mathbf{F}-\mathbf{C})}, \qquad (4.1)$$

where the subscript h represents the scale under consideration. Because of the linearity of the transform the score $\%ACC_h$ can be calculated in either the wavelet space or in the data space. If done in the data space, before inverse WTing is carried out the wavelet coefficients at all other scales

except h should first be set to 0. Later we show how this is done with some graphical examples.

The percent each scale contributes to MSE can be calculated by

$$\%MSE_h = \frac{\|A_h - F_h\|_f^2}{\|\mathbf{A} - \mathbf{F}\|_f^2}. \qquad (4.2)$$

The score $\%MSE_h$ can also be calculated in the wavelet space or in the data space (with the same recommendations about inverse transforming).

Figure 4.4 presents each scale of the analysis and forecast in the data space (alternatively, we could have presented these figures in the wavelet space). The solid contours are positive numbers, while the dashed are for negative numbers. These figures are prepared by WTing the analysis and forecast field and inverse WTing. For each scale h shown, before inverse WTing, all wavelet coefficients at scales other than h are first set to 0. Table 4.1 shows the values calculated for $\%ACC_h$ and $\%MSE_h$ (in the data space).

Examination of $\%ACC_h$ shows that about 80% of ACC is explained by scale $h = 1$. Simulations, like those carried out in Section 4.1, reveal this to be a general trait. Even fields which are widely different from each other produce large values for $\%ACC_1$. This phenomena occurs is because, in a sense, $\%ACC_1$ is attempting to measure the correlation between the means of the two fields which will, in meteorological forecast fields, always be somewhat close. Because $\%ACC_1$ is so large, little is left to explain with the other $\%ACC_h$. Thus we should judge the remaining $\%ACC_h$ relative to each other, not necessarily to $\%ACC_1$. Scales 2 and 3 explain the bulk of the remaining ACC after scale 1 is taken into consideration, while scales 4 and 5 add little. The same kinds of interpretations can be given for $\%MSE_h$.

The scores $\%ACC_h$ and $\%MSE_h$ are helpful in only a limited sense. They do give increased information about forecast performance by scale with respect to the common measures, but we can do better. To motivate how, reconsider the geometric interpretation of the closeness scores introduced in Section 2. Recall the correlation r can be interpreted as the cosine of the angle between two vectors (the ACC will not be used here). Likewise, RMSE can be interpreted as a function of a distance between these two vectors. Both of these measures are important to understand aspects of field closeness; both should be examined *simultaneously* for any given field pair. Consideration of either independent of the other ignores available information. A third measure, given below, is introduced that will complete the geometric picture. The geometric (and energy) interpretation will be emphasized in the construction of a closeness measure of the forecast-analysis fields by scale. The L_2 matrix norm will be used throughout, i.e. equation (2.2) and (2.3) RMSE and r will be calculated using norms (not, as is usual, by calculations with individual grid points). The mean of each matrix A_h and F_h is subtracted before any calculations are performed.

Each score, r and RMSE, can be calculated in the usual manner for each scale h. Designate these scores as r_h and $RMSE_h$. Because of linearity both can be calculated in the wavelet or data space. Values for $RMSE_h$ calculated in by space, however, will not be sum to RMSE. This is because, in the wavelet space, each matrix A_h and F_h consist of only n_h grid points where $\sum_h n_h = n$. In the data space each matrix A_h and F_h consist of n data points. Recall, in Section 2, that calculation of RMSE included $\sqrt{n}$ in the denominator. If we were to calculate the RMSE at each scale h in the wavelet space, each scale would have a different denominator

($\sqrt{n_h}$). When we calculate RMSE by scale in the data space the denominator equals $\sqrt{n}$. The two calculations are related by

$$\frac{\sqrt{n}}{\sqrt{n_h}} RMSE_h(data) = RMSE_h(wavelet). \tag{4.3}$$

Table 4.1 Comparing scores by percent and level for each scale

scale	1	2	3	4	5
$\%ACC_h$	0.80	0.09	0.10	0.01	0.00
$\%MSE_h$	0.48	0.16	0.20	0.08	0.08
r_h	0.998	0.968	0.978	0.808	0.409
$RMSE_h$	11.56	6.98	7.87	4.85	5.68
ER_h	0	0.01	0.25	0.03	0.51

Define the energy ratio (ER) by scale as

$$ER_h = |1 - \xi|, \tag{4.4}$$

where $\xi = \min\left\{\frac{\|\mathbf{A}_h\|_2^2}{\|\mathbf{F}_h\|_2^2}, \frac{\|\mathbf{F}_h\|_2^2}{\|\mathbf{A}_h\|_2^2}\right\}$. Defining ξ in this manner insures that ER is bounded by 0 and 1. The ER can be thought of, as was done in Section 3, by energy arguments. Each scale has certain amount of energy, and it is helpful to ask whether $\mathbf{F}$ has the same amount of energy for each scale h as does $\mathbf{A}$. The geometric motivation for the ER can be found in Fig. 2.1. The squared norm of each matrix $\mathbf{A}$ and $\mathbf{F}$ is representative of it's length. If the length of each vector at scale h is identical, then $ER_h = 0$. Of course, through the Law of Cosines (2.5) given in Section 2, it is possible to express ER as a function of r and MSE, but examining it independently can lead to important insights of field closeness. Table 4.1 lists the scores received for r_h, $RMSE_h$, and ER_h.

For the fields under consideration, ER_h indicates that at scale 1, **A** and **F** are, in the geometric sense, the same length, or hold equivalent amounts of energy. The correlation r_h, or angle between fields, also indicates only excellent agreement. The distance between the two fields, $RMSE_h$, however is fairly wide. The same interpretation for scale 2 as for scale 1 holds for ER_h and r_h, but here $RMSE_h$ reveals the fields to be somewhat closer. The ER_h for scale 3 shows that the two fields are now not the same length or the same energy, but that the angle between the fields is still small. The distance between fields, $RMSE_h$, is still modestly far. Scale 4 reveals that the energy or length of the two fields is nearly identical, but the angle is a bit wider, although the distance has closed somewhat. Scale 5 shows that the length or energy is widely dissimilar between the fields, the angle is also quite wide, although the distance is only moderate.

Examination of Fig. 4.4 will confirm, in a subjective way, most of the results of the scores. Strict visual interpretation is constrained by the fact that the L_2 space where these calculations are taking place is impossible to visualize. Additionally, one would not wish to examine images of forecast and analysis pairs by scale operationally. Instead, values of the about closeness measures r_h, $RMSE_h$, and ER_h could be kept track of, and particular exemplary or poor performance could be flagged and examined in detail (perhaps visually or with scores like $\%ACC_h$ and $\%MSE_h$). Use of a multivariate score will naturally lead to greater insight of forecast performance than will use of univariate scores.

5. CONCLUSION

Field forecast verification is a difficult problem, and one that is very often ignored. Typical users are content with applying familiar point scores, even with the knowledge that these scores are biased and non-robust. The high level of covariance between different grid points breaks the assumption of independence so that sampling distributions for scores like the RMSE are unknown. Without knowledge of a sampling distribution, it is impossible to gauge the significance of any particular score received and impossible to rate the difference between two competing forecasts.

The discrete wavelet transform is shown to have applications in univariate data analysis problems such as time series analysis and multivariate data analysis problems, such as image compression and detailed field closeness measures.

This article should not be viewed as an answer to the development of a robust sufficient statistic to measure field/image closeness. Rather, it is an attempt to incrementally improve interpretation of forecast verification measures that exist. It is clear the complexity and dimensionality of field forecast verification, as outline in Section 2, may be such that the formation of a perfect score, multi-or univariate, is impossible. This is why attempts to "break-up" the scoring system, such that we can examine closeness over a number of different scales, are helpful. The use of wavelets in partitioning a field is especially intriguing because the partitions can be made orthogonal, i.e. statistically independent. This work is only the first step in utilizing wavelets in this context and no doubt further improvements will suggest themselves in the future.

ACKNOWLEDGMENTS

We would like to thank David Ruppert and Wayne Bresky for there helpful comments along the way. Partial support for this research was provided by NIH training grant # ES07261. The remaining funding came from black market profits of illicit cigarette sales to renegade Canadians.

LIST OF FIGURES AND TABLES

Figure 4.1 A particular realization of the simulated forecast field. Contour levels are the same as Fig. 3.1a.

Figure 4.2 Mean results of 50 simulations for each value of s of (a) ACC and (b) RMSE. The vertical axis is the unit of the score. In each the horizontal axis is the variance parameter s.

Figure 4.3 The 36-hour NCAR CCM-2 forecast corresponding to Fig. 3.1a.

Figure 4.4 A series of images in the data space comparing the analysis and forecast fields at each of the five orthogonal wavelet scales.

BIBLIOGRAPHY

Anderson, T. W., 1984. *An Introduction to Multivariate Statistics*, Second Edition, Wiley, New York, 675 pp.

Branstator, G., A. Mai, and D. Baumhefner, 1993: Identification of highly predictable flow element for spatial filtering of medium- and extended-range numerical forecasts. *Monthly Weather Review*, **121**, 1786-1802.

Chas, B.F., and I. Naito, 1995: Wavelet analysis provides a new tool for studying Earth's rotation. *Eos*, **76**, 161-165.

Chui, C.K., L. Montefusco, and L. Puccio, 1994. *Wavelets: Theory, Algorithms, and Applications*, Academic Press, 627 pp.

Cressie, N., 1991. *Statistics for Spatial Data*, Wiley, New York, 900 pp.

Daley, R., 1993. *Atmospheric Data Analysis*, Cambridge University Press, 457 pp.

Daubechies, I., 1992. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 357 pp.

DeVore, R.A., 1993: Adaptive wavelet bases for image compression. In Laurent, P.J., A. Le Méhauté, and L.L. Shumaker, (eds.), *Wavelets, Images, and Surface Fitting*, A.K Peters, 197-219.

Donoho, D.L., 1992: De-noising by soft-thresholding. Technical Report 409, Stanford University.

Donoho, D.L., and I.M. Johnstone, 1994: Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425-455.

Donoho, D.L., and I.M. Johnstone, 1995: Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, **90**, 1200-1224.

Donoho, D.L., I.M. Johnstone, G. Kerkyacharian, and D. Picard, 1995: Wavelet shrinkage: asymptopia? (with discussion) *J. Royal Statistical Society*, **57**, 301-369.

Gao, W., and B.L. Li, 1993: Wavelet analysis of coherent structures at the atmosphere-forest interface. *J. Applied Meteorology*, **32**, 1717-1725.

Goel, P., and B. Vidakovic, 1995: Wavelet transformations as diversity enhancers. Discussion paper 95-04, ISDS, Duke University.

Golub, G.H., and C.F. Van Loan, 1989. *Matrix Computations*. Johns Hopkins Press, Baltimore, 642 pp.

Hagelberg, C. R. and N.K.K. Gamage, 1994: Applications of structure preserving wavelet decompositions to intermittent turbulence: a case study. In Foufoula-Georgiou, E., and P. Kumar, eds. *Wavelets in Geophysics*. Academic Press, New York, 45-80.

Hancock, M. S. and Wallis, J. R., 1994: An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields. *Journal of the American Statistical Association*, 89, 368-390.

Hoffman, R.N., Z. Liu, J.F. Louis, and C. Grassotti, 1995: Distortion representation of forecast errors. *Monthly Weather Review*, **123**, 2758-2770.

Johnstone, I.M., and B.W. Silverman, 1994: Wavelet threshold estimators for data with correlated noise. Technical Report, Stanford University.

Katul, G., and B. Vidakovic, 1995: The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering. Discussion paper 95-08, ISDS, Duke University.

Kumar, P., and E. Foufoula-Georgiou, 1993: A new look at rainfall fluctuations and scaling properties of spatial rainfall using orthogonal wavelets. *J. Applied Meteorology*, **32**, 209-222.

Kumar, P., and E. Foufoula-Georgiou, 1994: Wavelet analysis in geophysics: an introduction. In Foufoula-Georgiou, E., and P. Kumar (eds.), *Wavelets In Geophysics*, Academic Press, 373 pp.

Lau, K.M., and H. Weng, 1995: Climate signal detection using wavelet transform: how to make a time series sing. *BAMS*, **76**, 2391-2402.

Laurent, P.J., A. Le Méhauté, and L.L. Shumaker, 1993. *Wavelets, Images, and Surface Fitting*, A.K Peters, 528 pp.

Livezey, R.E., J.D. Hoopingarner, and J. Huang, 1995: Verification of official monthly mean 700-hPa height forecasts: an update. *Weather and Forecasting*, **10**, 512-527.

Meyers, S.D., B.G. Kelly, and J.J. O'Brien, 1993: An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of Yanai waves. *Monthly Weather Review*, **121**, 2858-2866.

Miyakoda, K., G.D. Hembree, R.F. Strickler, and I. Shulman, 1972: Cumulative results of extended forecast experiments I. Model performance for winter cases. *Monthly Weather Review*, **100**, 836-854.

Murphy, A.H., and E.S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Monthly Weather Review*, **117**, 572-581.

Murphy, A.H., 1991: Forecast verification: its complexity and dimensionality. *Monthly Weather Review*, **119**, 1590-1601.

Nason, G.P., 1994: Wavelet regression by cross-validation. Technical Report 447, Stanford University.

Perrie, W. and Toulany, B., 1989: Correlations of Sea Level Pressure Fields for Objective Analysis. *Monthly Weather Review*, 117, 572-581.

Preisendorfer, R.W., 1988. *Principal Component Analysis in Meteorology and Oceanography*. Elsevien, New York, 425 pp.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, 1992. *Numerical Recipes in C*, Cambridge University Press, Second Edition, 994 pp.

Radok, U., and T.J. Brown, 1993: Anomaly correlation and an alternative: partial correlation. *Monthly Weather Review*, **121**, 1269-1271.

Saito, N., 1994: Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length comparison. In Foufoula-Georgiou, E., and P. Kumar (eds.), *Wavelets In Geophysics*, Academic Press, 373 pp.

Sampson, P. D. and Guttorp, P., 1992: Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *Journal of the American Statistical Association*, 87, 108-119.

Serrano, E., R. Compagnucci, and M. Fabio, 1992: The use of wavelet transform for climatic estimates. *Proceedings Fifth Int. Meeting on Statistical Climatology*, Toronto, Canada, AES, Environment Canada, 259-262.

Snedecor, G. W. and Cochran, W. G., 1989. *Statistical Methods*, Eighth Edition, Iowa State University Press, Iowa, 503 pp.

Taylor, C.C., 1991: Measure of similarity between two images. In A. Possolo (ed.), *Spatial Statistics and Imaging*, IMS Lecture Notes, Vol. 20, 426 pp.

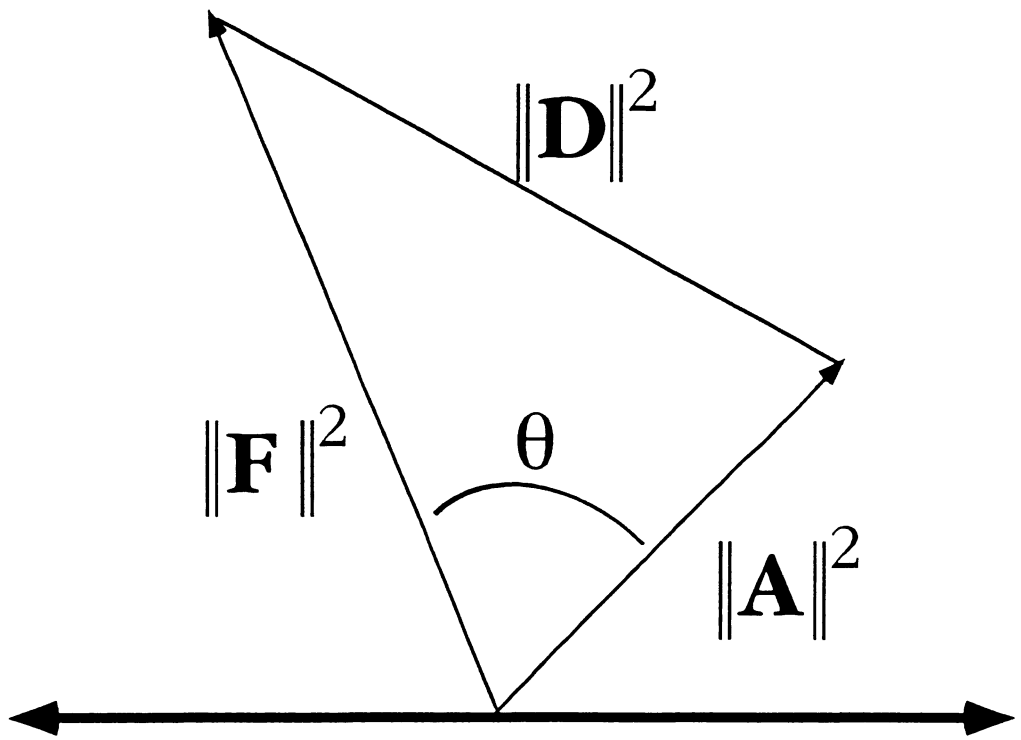
Van Den Dool, H.M., and L. Rukhovets, 1994: On the weights for an ensemble-averaged 6-10-day forecast. *Weather and Forecasting*, **9**, 457-465.

Vidakovic, B., and P. Müller, 1994: Wavelets for kids. Discussion Paper 94-13, Duke University.

Wallace, J.M., and D.S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review*, **109**, 784-812.

Weng, H., and K.M. Lau, 1994: Wavelets, period doubling, and time-frequency localization with application to organization of convection over the tropical Western Pacific. *J. Atmospheric Sciences*, **51**, 2523-2541.

Wilks, D. S., 1995. *Introduction to Statistical Method in the Atmospheric Sciences*, Academic Press, New York, 467 pp.



$$\|\mathbf{D}\|^2 = \|\mathbf{A}\|^2 + \|\mathbf{F}\|^2 - 2\|\mathbf{A}\|\|\mathbf{F}\|\cos\theta$$

$$r = \cos\theta$$

$$\text{nMSE} = \|\mathbf{D}\|^2 = \|\mathbf{A} - \mathbf{F}\|^2$$

Figure 2.1

(see Persson)
REF on p. 10

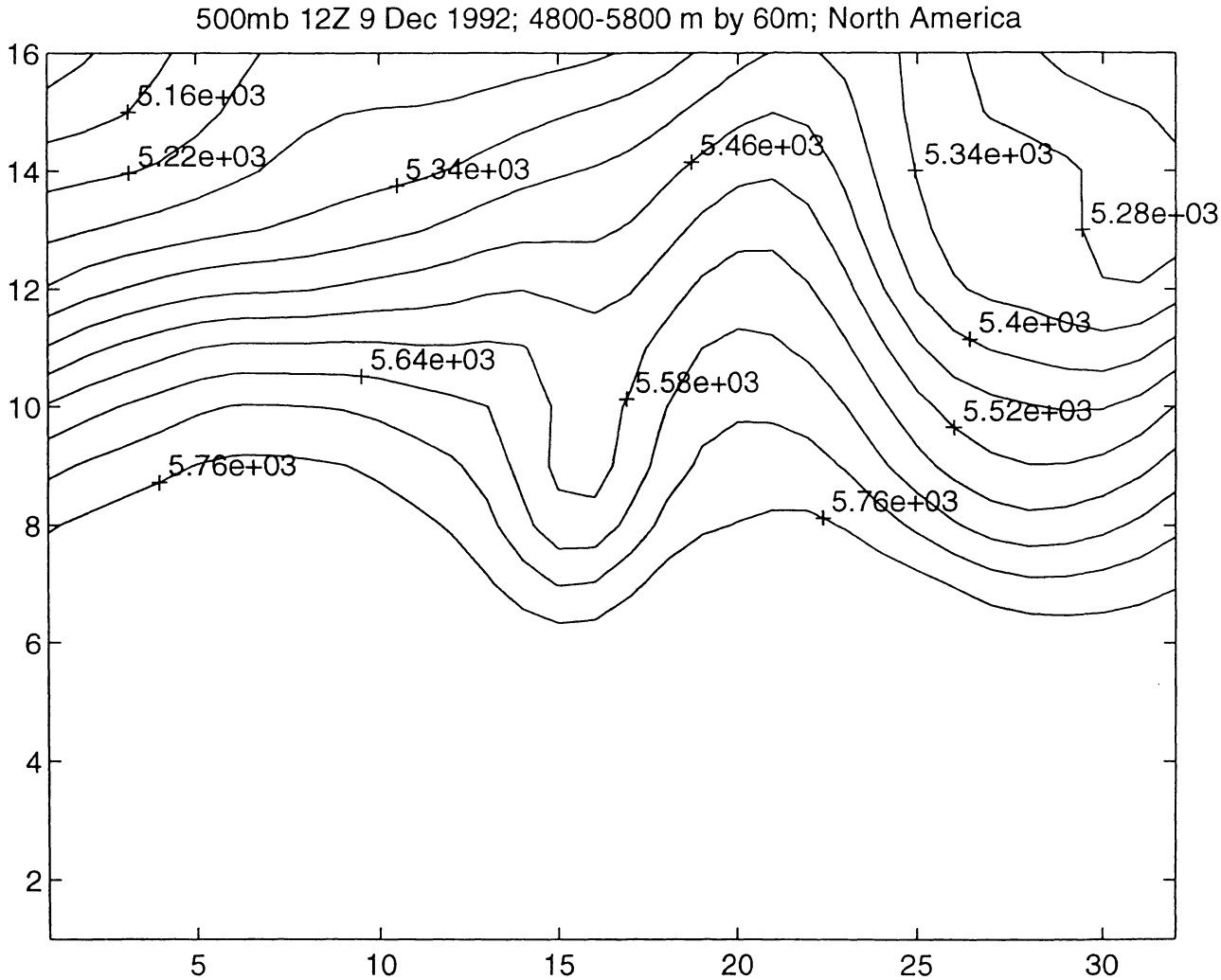

500mb 12Z 9 Dec 1992; 4800-5800 m by 60m; North America
5.16e+03
5.22e+03
5.34e+03
5.46e+03
5.34e+03
5.28e+03
5.4e+03
5.64e+03
5.58e+03
5.52e+03
5.76e+03
5.76e+03
16
14
12
10
8
6
4
2
5
10
15
20
25
30

Figure 3.1a

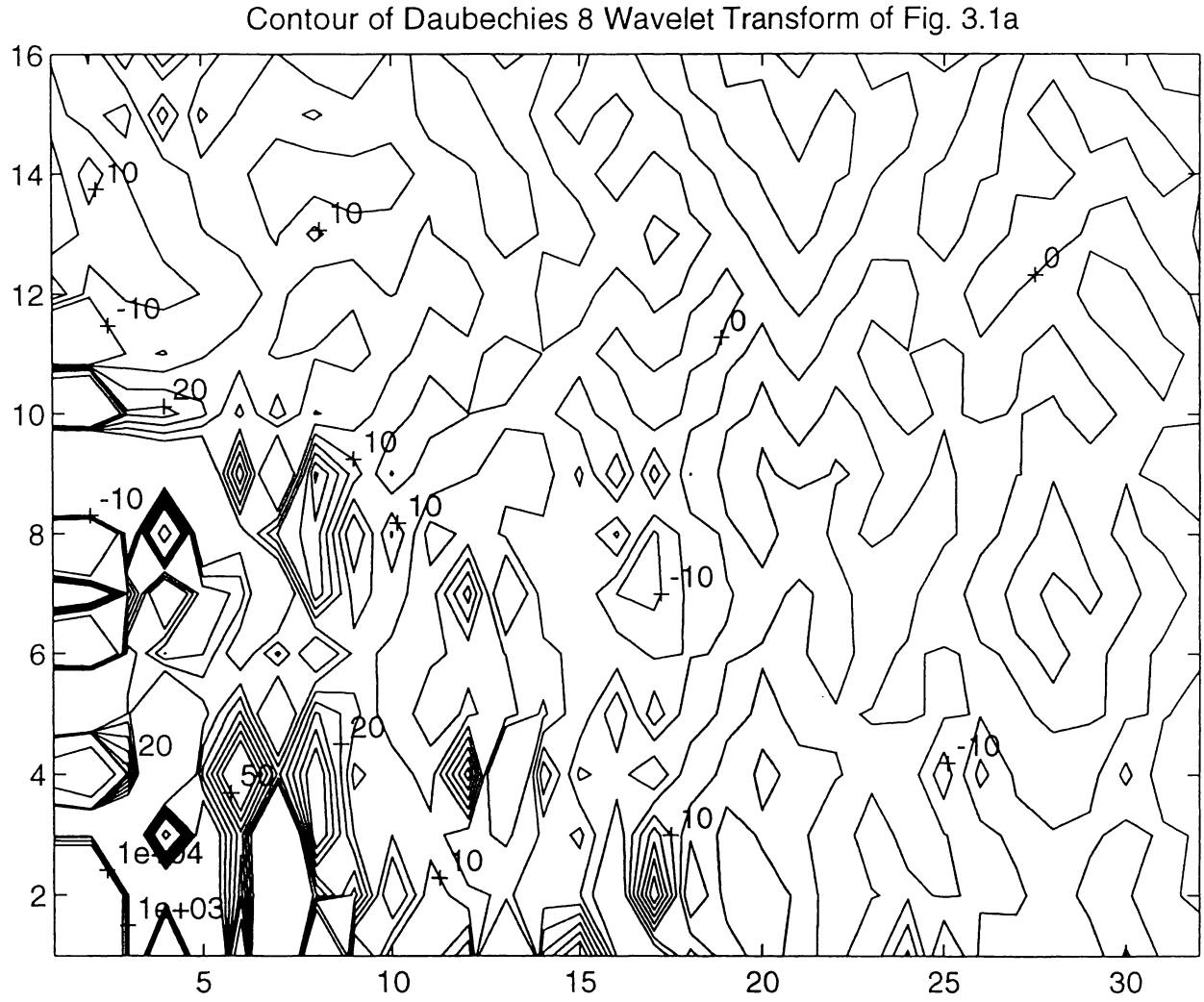

Contour of Daubechies 8 Wavelet Transform of Fig. 3.1a
16
14
12
10
8
6
4
2
5
10
15
20
25
30
10
10
-10
20
10
10
-10
20
50
1e+04
1e+03
10
-10
0
0
-10
10
10

Figure 3.1b

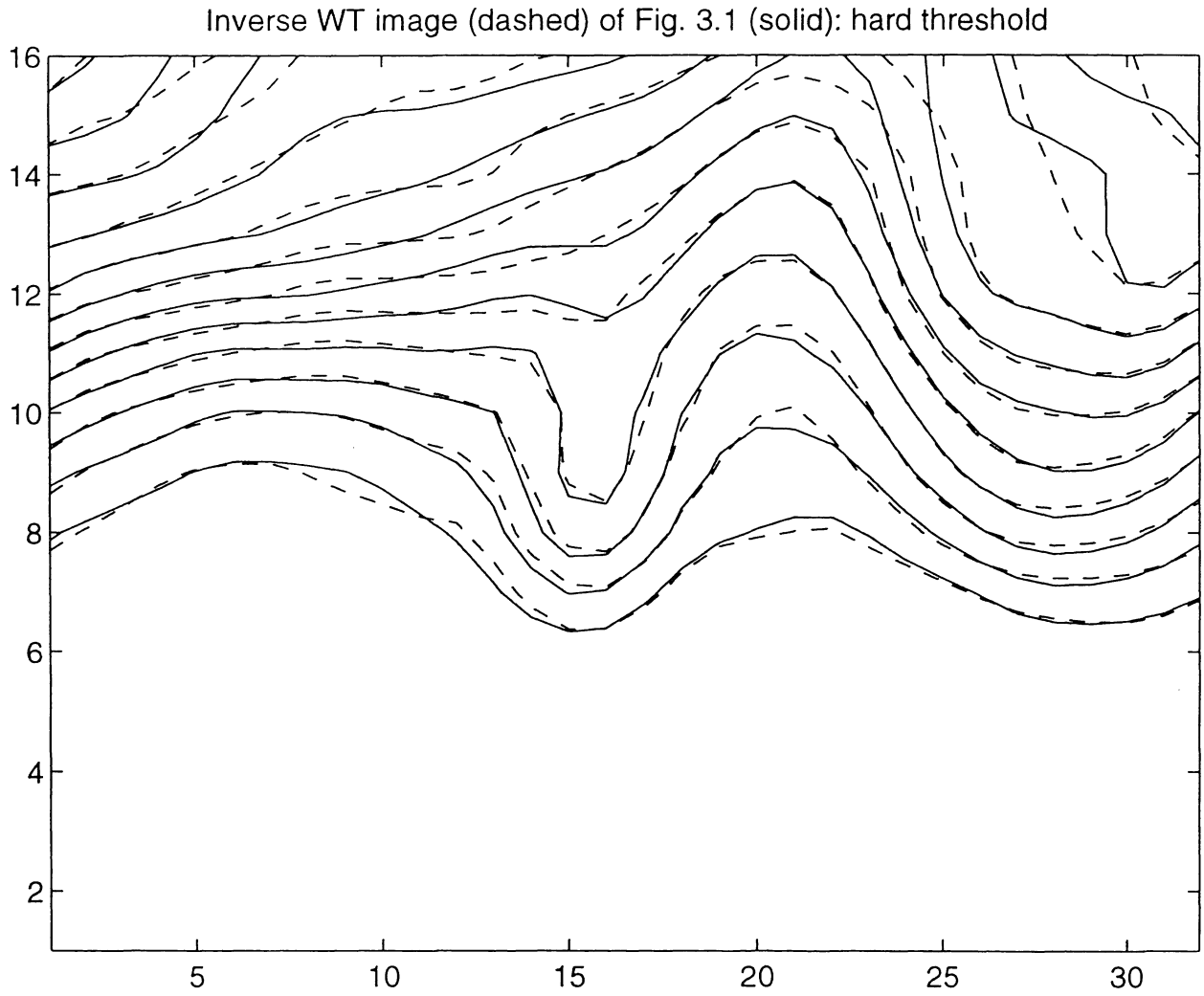

Inverse WT image (dashed) of Fig. 3.1 (solid): hard threshold
16
14
12
10
8
6
4
2
5
10
15
20
25
30

Figure 3.2

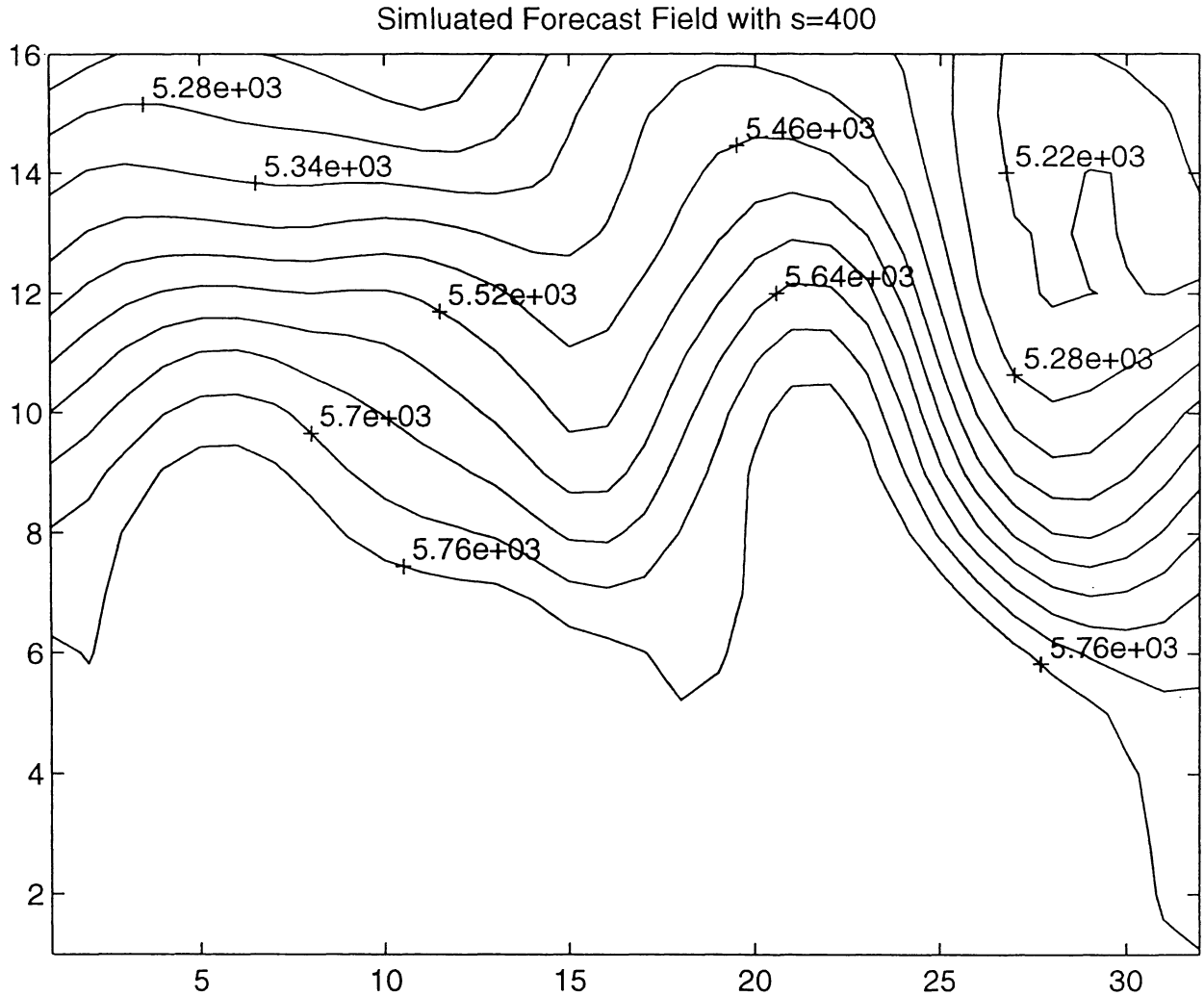

Simluated Forecast Field with s=400
5.28e+03
5.34e+03
5.46e+03
5.22e+03
5.52e+03
5.64e+03
5.28e+03
5.7e+03
5.76e+03
5.76e+03
16
14
12
10
8
6
4
2
5
10
15
20
25
30

Figure 4.1

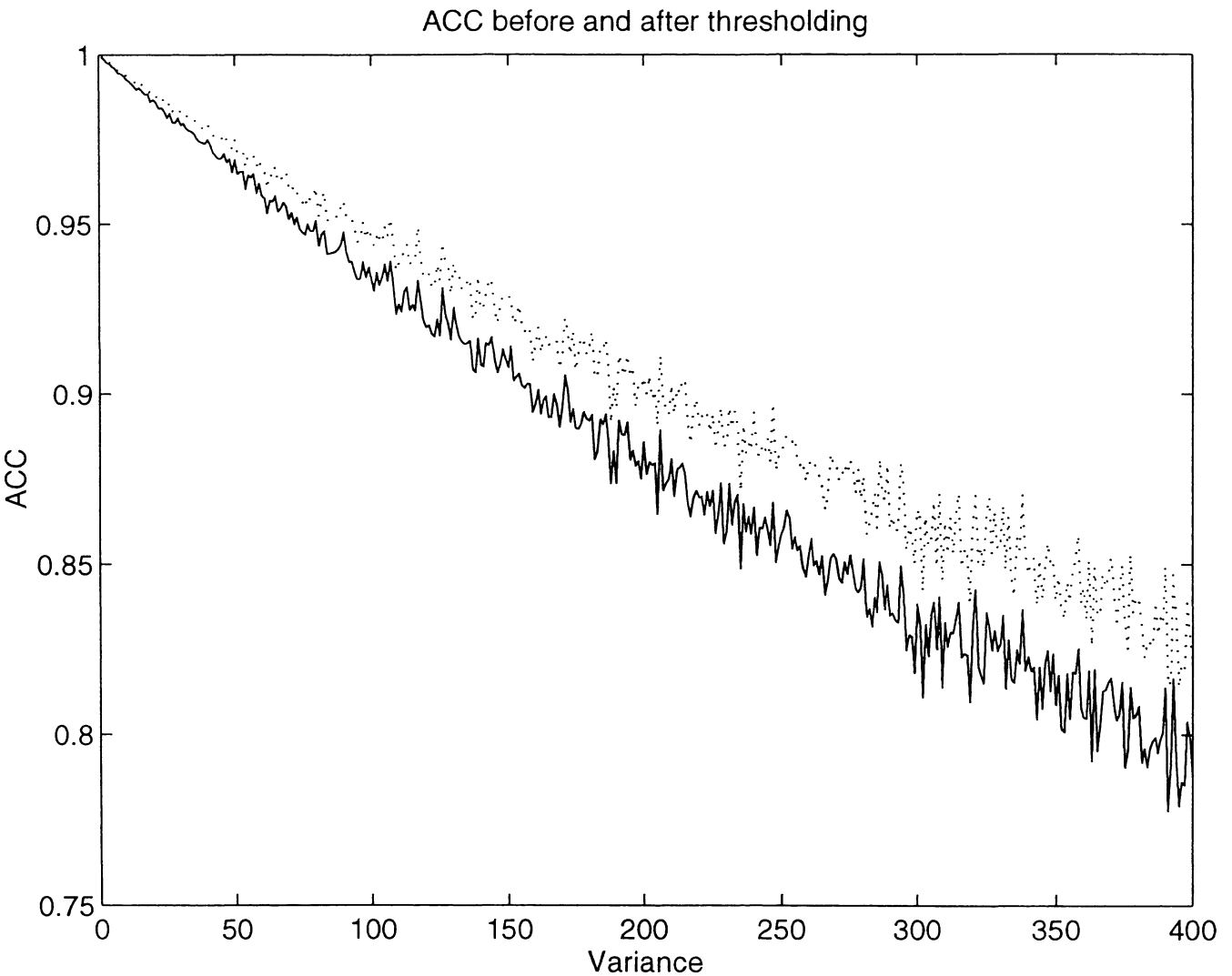

ACC before and after thresholding
ACC
Variance
1
0.95
0.9
0.85
0.8
0.75
0
50
100
150
200
250
300
350
400

Figure 4.2a

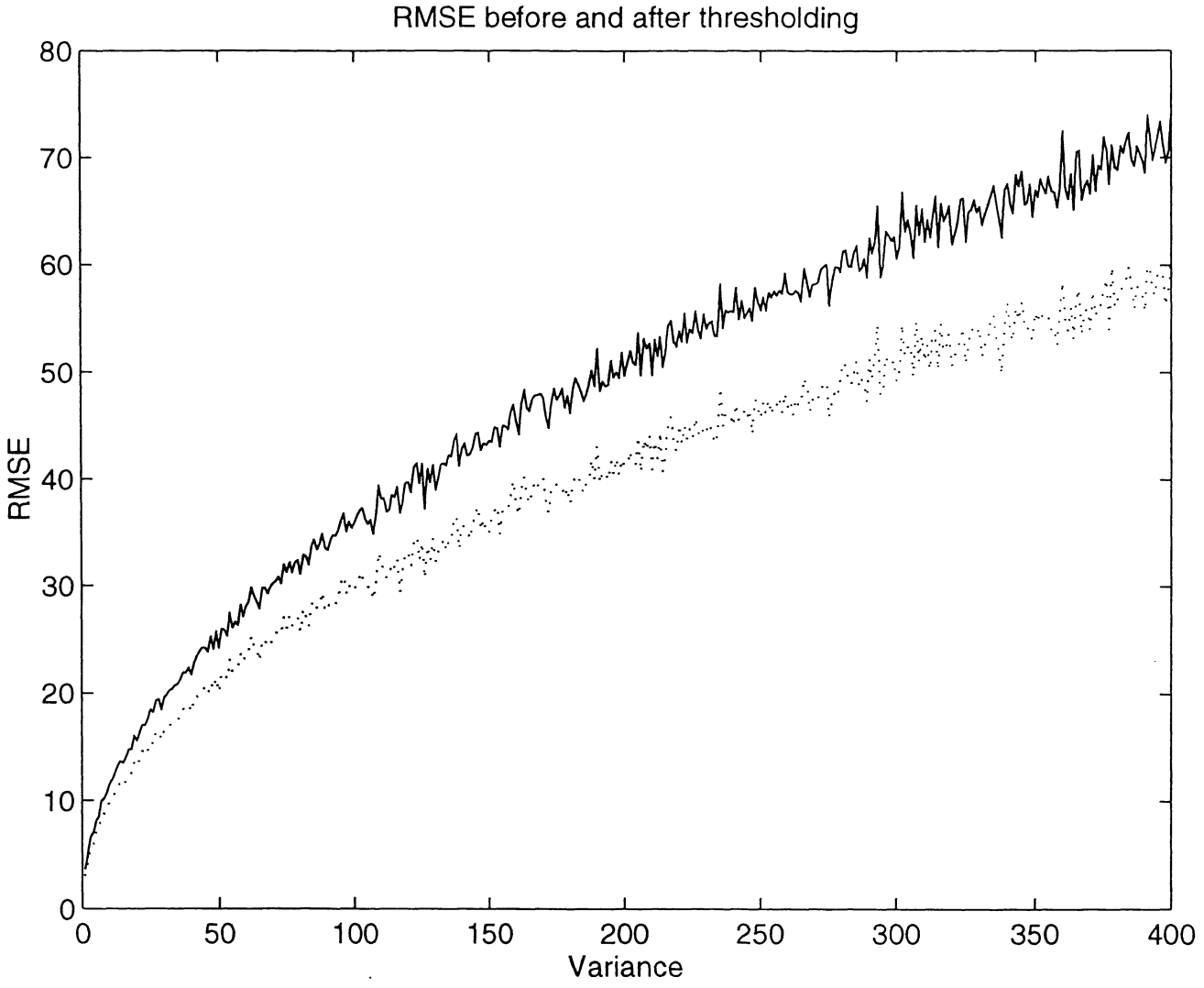

RMSE before and after thresholding
RMSE
Variance
0
10
20
30
40
50
60
70
80
0
50
100
150
200
250
300
350
400

Figure 4.2b

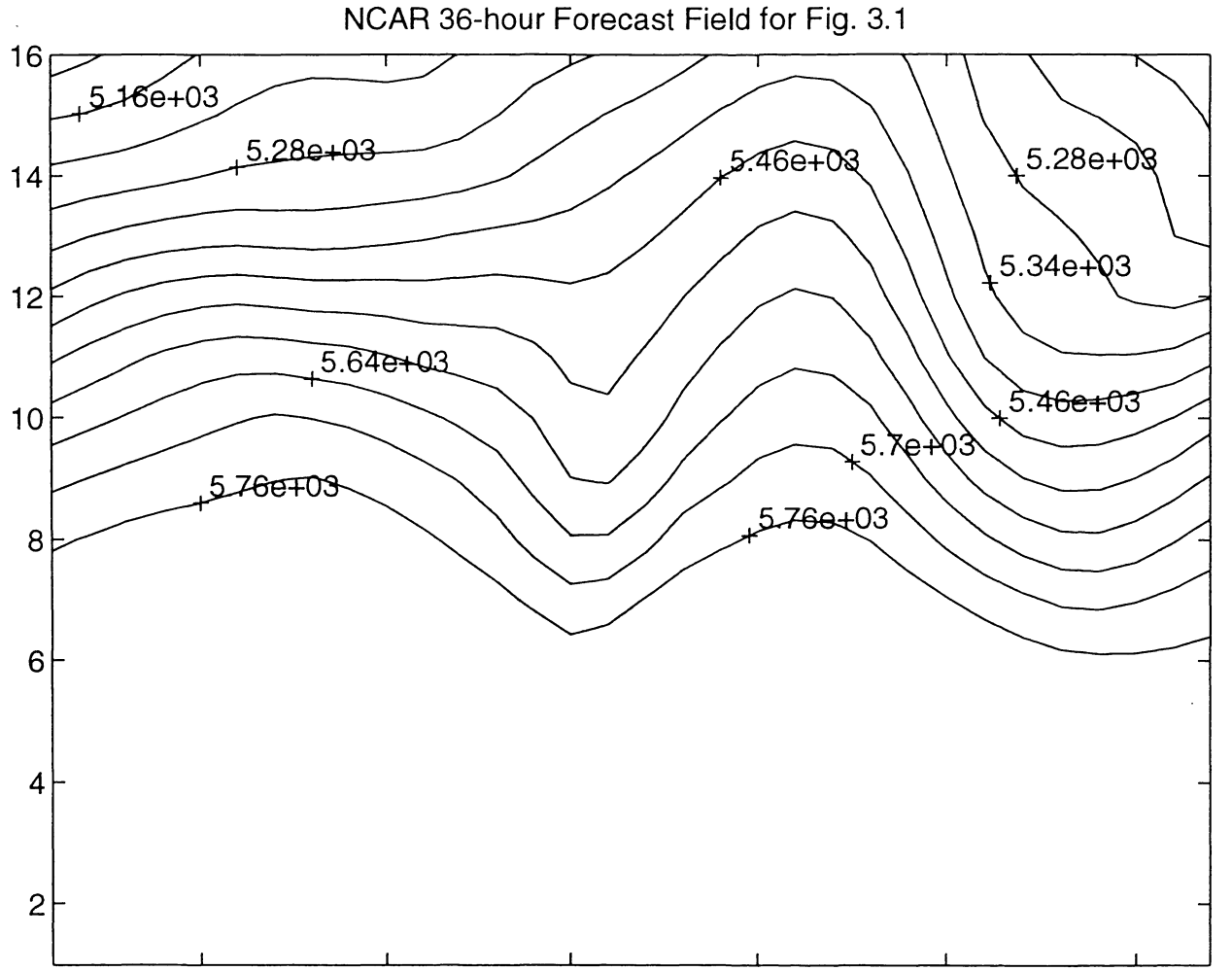

NCAR 36-hour Forecast Field for Fig. 3.1
5.16e+03
5.28e+03
5.46e+03
5.28e+03
5.34e+03
5.64e+03
5.46e+03
5.7e+03
5.76e+03
5.76e+03
16
14
12
10
8
6
4
2
5
10
15
20
25
30

Figure 4.3

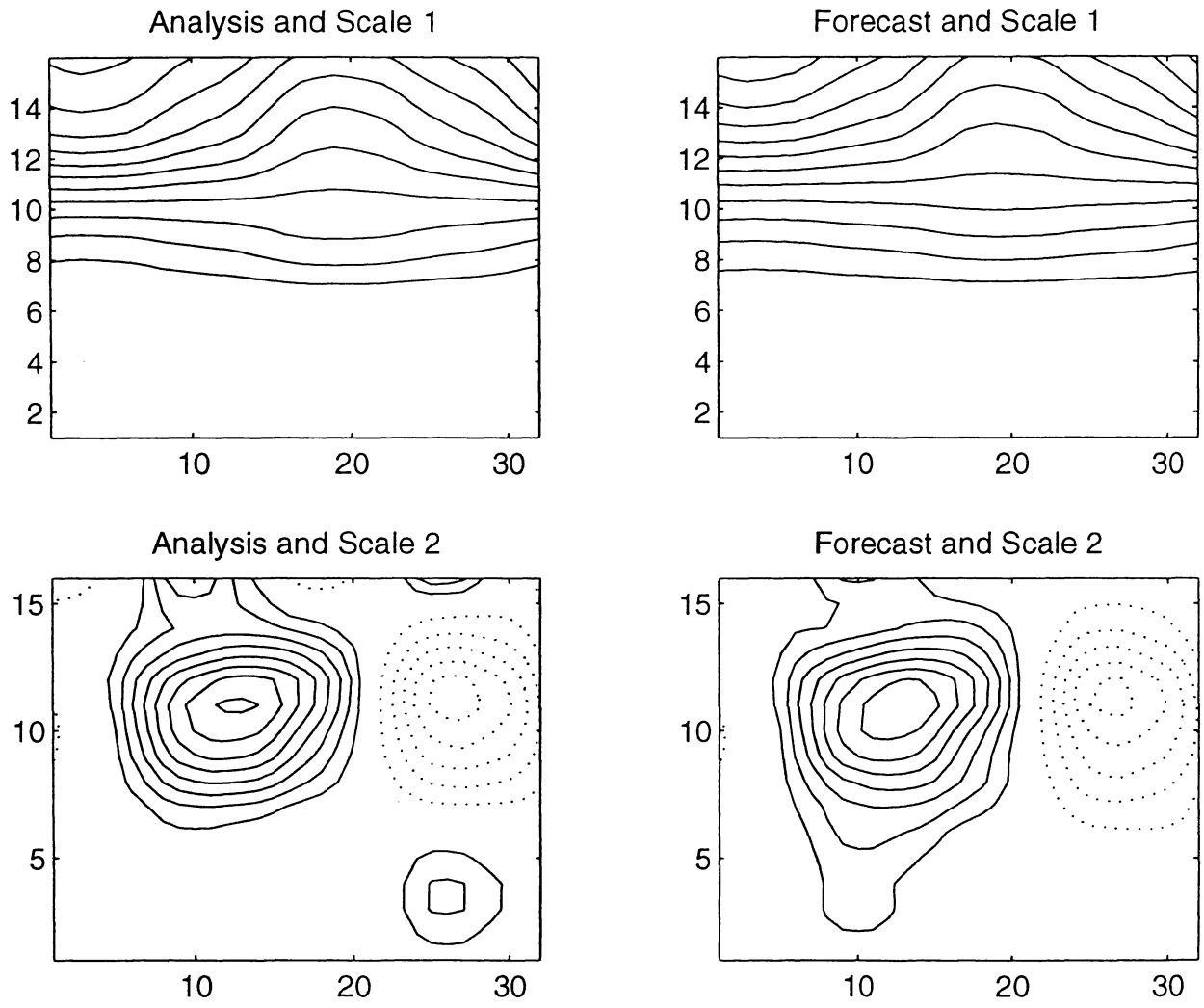

Analysis and Scale 1
Forecast and Scale 1
Analysis and Scale 2
Forecast and Scale 2
14
12
10
8
6
4
2
15
5
10
20
30

Figure 4.4

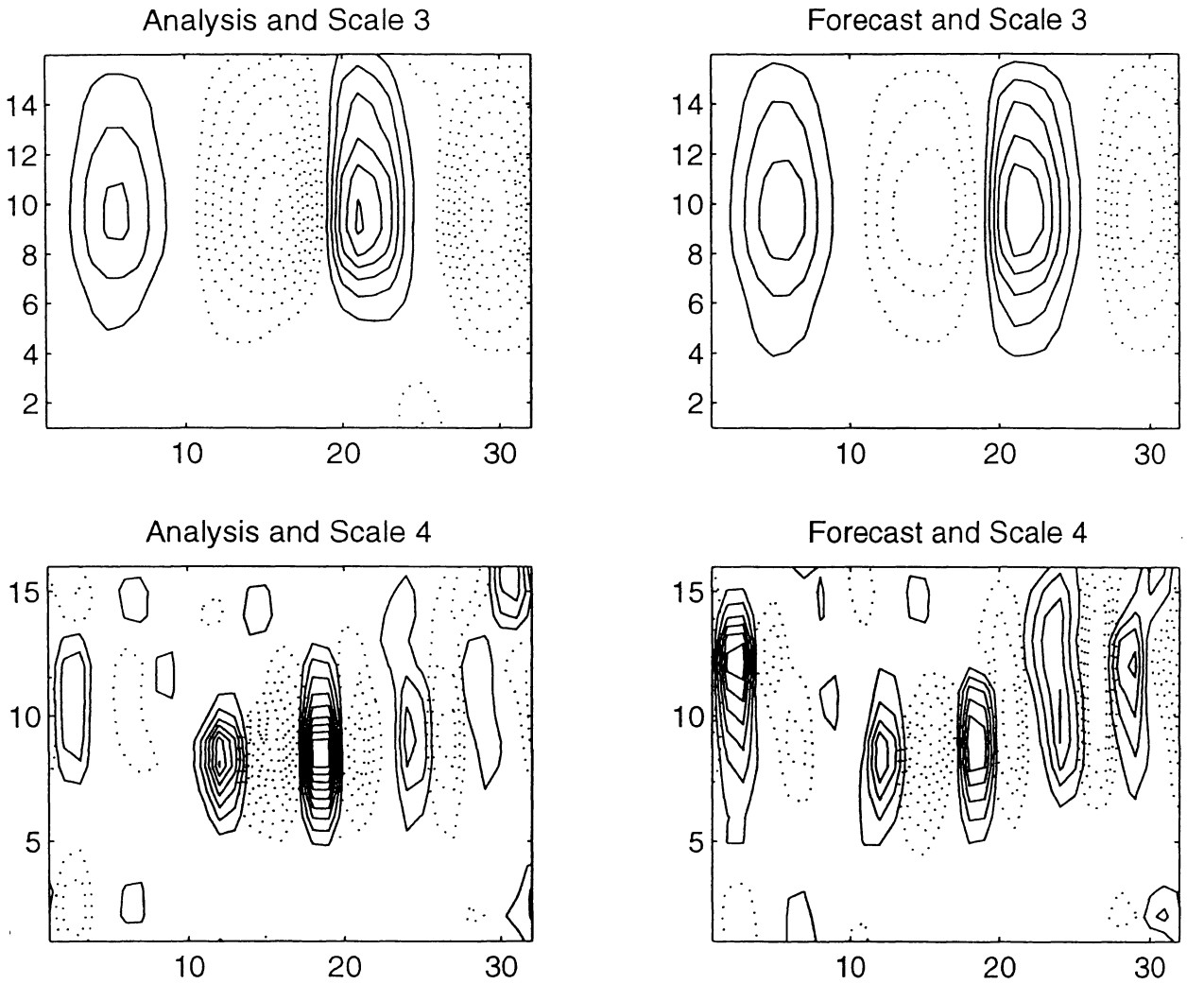


Figure 4.4 (continued)

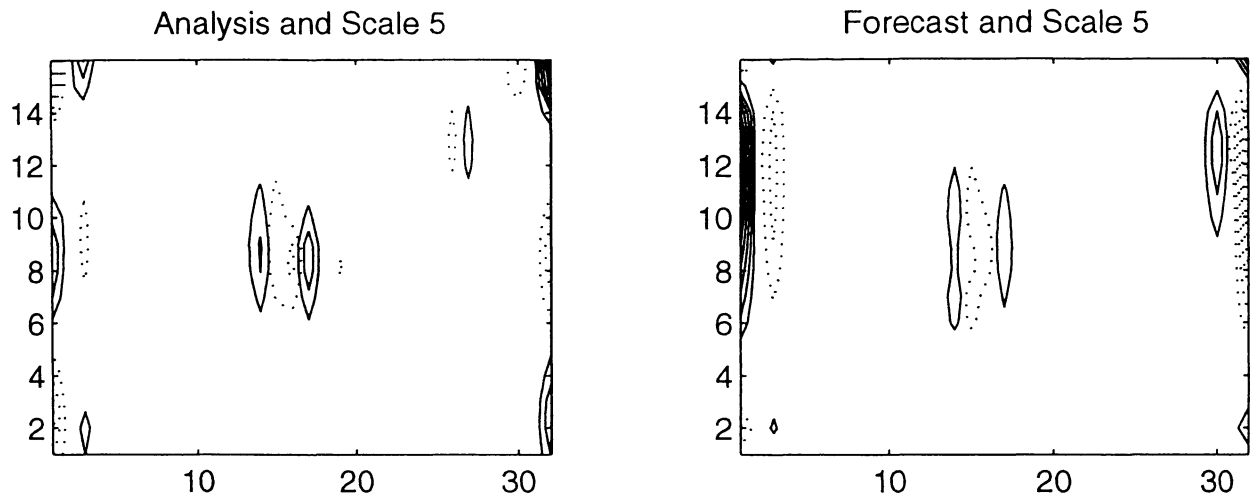


Figure 4.4 (continued)